

少数例にもとづく精密検定 (2 標本問題)

川崎医科大学 数学教室

仮 谷 太 一

(昭和61年9月10日受理)

Exact tests based on small samples

(Two sample problem)

Taichi KARIYA

Department of Mathematics,

Kawasaki Medical School,

Kurashiki 701-01, Japan

(Received on Sept. 10, 1986)

概 要

連続な確率変量データに基づく統計的仮説の検定においては、母集団分布の連続性を仮定し、同順位データはないものとして理論的展開を行っているが、実際問題としては同順位データの得られることが少ない。医学・生物学の実験では標本のサイズはあまり大きくないのが普通であり、その上、同順位データがあるとなると、標本サイズが十分大きいときの漸近分布に依存する検定手順は、甚だしく妥当性を欠くことになる。少数例でしかも同順位データがある場合、その検定統計量の精密分布は、サイズ無限大のときの漸近分布とは、かなり大きく食違うからである。

本論文では、少数例にもとづく精密検定を、Fisherの並べかえ検定・Wilcoxonの検定(独立2標本)、符号つき差の計による検定・Wilcoxonの符号順位検定(対応2標本)について、同順位データの存在を許す立場から考察し、そのコンピュータプログラムを提示して、これら検定法の適用を正確かつ容易にした。

Abstract

For testing some statistical hypothesis we will postulate the continuity of population distribution and derive the distribution of test statistic on the assumption that there are no ties. As a matter of fact, however, ties are observed frequently. In the experimental studies in medicine and biology we cannot expect the large sample. Moreover ties are observed frequently and then the exact distribution of test statistic will be discrepant from its asymptotic distribution as the sample size tends to ∞ . Accordingly we cannot test the hypothesis by using the asymptotic distribution in these cases.

In the present paper we discuss the exact non-parametric tests which are excellent theoretically and easy to understand intuitively, for example Fisher's permutation test and Wilcoxon's signed rank test. The computer programs for these exact tests are shown with the graphic output.

1. 序 説

臨床試験における症例数はもちろんのこと、動物実験の場合にも、標本の大きさにはおのずから限度がある。臨床試験において症例数を増やそうとすれば、確率標本としての random 性・比較可能性などの大切な性質が失われるのが普通である。動物実験の場合にも、ほぼ同様なことが言えるであろう。動物愛護の立場からも、実験動物の数を必要最小限にとどめるべきは当然のことであるから、医学・生物学の実験においては、常に少数例の制約を逃れることはできない。

医学・生物学の研究において、統計学が最も多く用いられるのは、なんらかの意味での科学的な仮説を検定するときである。一般に実験科学の研究においては、(1) まず仮説を立て、それを検証するために、(2) 実験を計画し (3) 実施し、そして最後に (4) 得られたデータにもとづき当初の仮説の正否を検定することが必要である。殊に、生体に関する実験結果には、それほど明白な傾向性・法則性を見出すことの困難な場合が多い。いくつかの考えられる仮説の中から、ある決定を下すためには、統計的な検定が必要になるというわけである。

生体における実験では、ある一つの特性値の変化に的をしほっても、それに関連する要因が多く、関連があるのかないのか分っていない要因もあり、さらに実験要因による変動に拮抗する個体差のウエイトが非常に大きい。そこで実験群と対照群、あるいは2つの処理群が、実験要因以外の要因については、均等になるようにすることが重要である。こうして2つの処理群を比較可能にするような実験計画の下で、実験を行い、その実験結果を比較検討するいわゆる2標本問題が基本となる。

ところで、あまり大きくないサイズの2標本にもとづいて、統計的仮説の検定を行おうとすれば、平均値の差・中央値の差、あるいは順位平均の差とかの検定統計量の分布を精密に定めることが必要になる。従来、統計量の分布については、標本サイズが十分大きくなったときの漸近分布を用い、それにもとづいて当該標本と同等かそれ以上に起りにくい標本を得る確率(有意確率またはP-値)を計算するのが普通であった。

しかし、観測された特性値がどのような分布型に従うかは、よく分らないのが通例であり、標本サイズも大きくない場合には、変量の分布型を特定しない統計的検定で、しかも漸近分布を用いない精密分布によるのが望ましいことは申すまでもない。

この論文では、以上の観点から、変量の分布型を特定しない統計的検定で、理論的にも優れており、直観的に分り易いものを取りあげる。そして任意の標本サイズに対応できる、検定統計量の精密分布を求めるコンピュータ・プログラムを提示して、容易に統計的検定が行えるようにした。提示したプログラムは、BASIC プログラムで、FM16 β 用に開発したものであるが、プログラムは極めて基本的なものであるから、容易に解説でき、他の機種用書き換えることができるであろう。

なお、少数例とは何例までかとの問いに答えることは簡単ではないが、漸近分布に頼れない範囲の大きさとでも言ったらよいであろうか。両標本の大きさが違い、同順位データ (ties)

が多いとき、漸近分布と精密分布の食い違いは大きくなるが、このプログラムでは、計算処理時間ともならみ合せて、一応18としている。インタープリター方式でなく、コンパイラー方式が利用できるとか、大型機にかけられるような場合には、20前後としてよいであろう。

2. 採りあげる統計的検定と数値例

ここでは、理論的にも優れており、しかも直観的に分り易い統計的検定として、次のものを取りあげることにする。

A. 独立な2標本にもとづく non-parametric 検定

(1) Fisher の並べかえ検定

……平均値の差 またはずれの検定……

(2) Wilcoxon の精密検定

……中央値の差 またはずれの検定……

B. 対応する2標本にもとづく non-parametric 検定

(1) 符号つき差の計による検定

……中央値（平均値）の差の検定……

(2) Wilcoxon の符号順位和精密検定

……中央値の差の検定……

A. B.とも、2つの母集団分布の連続性だけを仮定する。特にB.では実験要因以外の要因による誤差は標本内・標本間ともに独立であるよう、うまく管理設計されているものとする。分布の連続性を仮定するとき、理論の上では観測値はすべて異なる値を取り、同一の値（データの大きさの順位で考えるときは同順位データ (tie)）になることはないはずである。しかし実際の観測データには、tieの表われることは、それほど珍しいことではない。tiesの存在は、漸近分布を用いる場合には大きな障害となるが、精密分布にもとづく検定の場合には、特に問題にはならない。

次に、A. B.の数値例を掲げて、問題の輪郭を具体的なものにしておこう。

A. の数値例

尿量のほぼ等しい腎臓病の患者11名を、管理可能な要因について層に分け、同一の層の患者に対して薬剤P, Qを無作為に割り付けた。そうして各人につき1週間の尿量を測って、次表のデータを得た。（表の数値は、各人の尿の総量を100で割り、さらにその値から100を引いた値である。）

薬剤P, Qによって尿量に差があると言えるか？

帰無仮説：薬剤P, Qにより尿量に差はない。

表1 P, Q剤投与群における尿量

	標本の大きさ	データ
P剤投与群(X)	5	31 28 19 31 33
Q剤投与群(Y)	6	40 33 47 23 37 38

B.の数値例

15人の健常者に対し、ある薬物を注射する直前と注射20分後に、耳たぶから採取した血液についてその塗抹標本をつくり、白血球500個中の好酸球の百分率を調べたところ、つぎのデータが得られた(数値は%を示す)。

—資料は柳川〔2〕103ページより—

表2 注射前後における白血球中の好酸球率(%)

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
前(X ₁)	2.0	10.2	6.0	7.8	2.2	3.4	1.6	4.2	6.6	7.6	3.8	8.6	4.2	11.2	2.0
後(X ₂)	3.4	15.6	5.4	10.8	5.6	6.2	3.8	3.4	5.2	10.0	2.0	13.2	5.4	13.6	4.4

薬物の注射によって、好酸球の百分率は、増加したといえるか?

帰無仮説: 注射によって好酸球の百分率は変らない(百分率の増加は見られない)。

3. 検定モデルとその理論(その1)

A. (1) Fisherの並べかえ検定

連続な母集団分布 $F(x)$, $G(y)$ からの、大きさ n_1 , n_2 の確率標本を、それぞれ X_1, X_2, \dots, X_{n_1} , Y_1, Y_2, \dots, Y_{n_2} とする。

このとき検定すべき帰無仮説は、 $H_0: F(t) = G(t)$ と表すことができる。これに対する対立仮説としては、もし実験を始める前に、Q剤の方がP剤よりも尿量をいっそう多くするという何等かの根拠があったとすれば、 $H_1: F(t) > G(t)$ とすることができる。両剤の優劣について特別のよりどころがなく、単に両剤の間に差があるかないかを検定したいときには、 $H_2: F(t) > G(t)$, または $F(t) < G(t)$ とするのが適当であろう。ここでは H_1 の場合について議論を進める。($F(t) < G(t)$ を対立仮説とするときには、PとQとを入れ替えて議論すればよく、また H_2 の場合の有意確率(P-値)は、 $F(t) > G(t)$ の場合のP-値と $F(t) < G(t)$ の場合のP-値の和として計算することができる)

いま、 X_1, X_2, \dots, X_{n_1} および Y_1, Y_2, \dots, Y_{n_2} の実際の観測値を、小文字 x_1, x_2, \dots, x_{n_1} ,

y_1, y_2, \dots, y_{n_2} で表すことにする。このとき Fisher の並べかえ検定の考え方は次のとおりである。

$X_1, \dots, X_{n_1}; Y_1, \dots, Y_{n_2}$ の母集団分布 $F(x), G(y)$ については、単に連続性だけが仮定されており、分布型などについて一切仮定をおいていないのであるから、データからの情報を用いて、どのような値が実現し易いとか、実現し難いとかを定量的に推測することは無理である。そこで、観測された実現値 $n_1 + n_2 = n$ 個を固定し、一緒にしてから、これらが現実の X 標本データ、 Y 標本データに分割されている＝その分割のされ方が、帰無仮説 H_0 の下で、ごく普通のものか、稀にしか起りそうもない分れ方かを判定しようとするのである。

H_0 の下では、 X, Y の母集団分布は全く同一であるから、 n 個のデータを、あらゆる順列に並べかえ、その前半の n_1 個が X 標本データ、残りの n_2 個が Y 標本データであるとすることは、すべて同様に確からしいことである。 n 個のデータの並べかえ方の総数は $n!$ 通りあるから (ties があってもごくわずかずつ異なると考える)、このことは、 n 次元空間における $n!$ 個の点集合をベースに考えることになる。この $n!$ 通りの場合に、 $D = \bar{X} - \bar{Y}$ (X 標本データの平均値と Y 標本データの平均値との差) を計算し、その確率分布を求めて、実際に観察された標本の平均値差が、ごく普通のものか、稀にしか起らないものかを、有意確率 (P -値) により判定しようとするものである。

この検定統計量 $D = \bar{X} - \bar{Y}$ の場合には、各標本内におけるデータの順番が変わっても平均値には無関係なので、 n 個のデータのすべての順列 $n!$ 通りについてではなく、 n 個の中から n_1 個を選ぶすべての組合せ ${}_nC_{n_1}$ 通りについて、 D の値を計算しその確率分布を求めればよい。このことは、平均値の差の代わりに、中央値の差を検定統計量とする場合にも同様である。

〔並べかえ検定の原理〕

並べかえ検定の考え方を明確にするために、きわめて簡単な例について説明しよう。

$$n_1 = 3; \quad x_1 = 1, \quad x_2 = 2, \quad x_3 = 3$$

$$n_2 = 2; \quad y_1 = 4, \quad y_2 = 5$$

としよう。このとき $n = n_1 + n_2 = 5$ である。

5 個のデータを 1 列に並べる並べ方の総数は、 $5! = 120$ であり、それぞれの順列の初めの 3 個のデータの平均値から、残りの 2 個のデータの平均値を引いた差 D の分布を求めるのであるが、これを次のように考えることができる。

5 個のデータから $n_1 = 3$ 個をとり出す取出し方の総数は、5 個のものから 3 個をとる組合せの数に等しく、 ${}_5C_3 = 5! / (3! 2!) = 10$ で、次のリストのとおりである。

No.	Xのデータ	Yのデータ	$D = \bar{X} - \bar{Y}$
1	1, 2, 3	4, 5	-2.5000
2	1, 2, 4	3, 5	-1.6667
3	1, 2, 5	3, 4	-0.8333

4	1, 3, 4	2, 5	-0.8333
5	1, 3, 5	2, 4	0
6	1, 4, 5	2, 3	0.8333
7	2, 3, 4	1, 5	0
8	2, 3, 5	1, 4	0.8333
9	2, 4, 5	1, 3	1.6667
10	3, 4, 5	1, 2	2.5000

この組合せの任意の一つ、例えばNo. 8 (2, 3, 5) (1, 4) をとりあげて、X 標本データ内、Y 標本データ内で順番を入れかえる仕方を考えてみると、(2, 3, 5) では $3! = 6$ 通り、その各々に対して (1, 4) では $2! = 2$ 通りあるから、都合 $3! \times 2! = 12$ 通りになる。他の 9 通りの組合せについても全く同様である。

10通りの組合せから、それぞれ12通りの順列が得られ、全部で120通り(=5個のデータ列の並べかえ順列の総数)が得られる。このことは

$${}_5C_3 \times (3!2!) = \frac{5!}{3!2!} \times (3!2!) = 5!$$

$$\text{一般に } {}_nC_{n_1} \times (n_1!n_2!) = \frac{n!}{n_1!n_2!} \times (n_1!n_2!) = n!$$

からも容易に理解される。

さてNo. 8 に対する平均値の差 D は $0.8333\cdots$ に等しく、(2, 3, 5) 内、(1, 4) 内で順番を入れかえて得られる12通りの場合についても、すべて $D = 0.8333\cdots$ である。よって、5 個のデータのすべての並べかえ120通りに対する平均値の差 D の分布を求めるには、5 個から3 個をとり出す組合せ ${}_5C_3 = 10$ 通りに対する平均値の差 D の分布を求めればよいことが分かる。

この場合の D の確率分布は次表のとおりである。確率欄の下段の(数字)は、すべてのデータの並べかえを基礎としたときの計算を示している。

D の値	-2.5	-1.66...	-0.833...	0	0.833...	1.66...	2.5	計
確 率	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{2}{10}$	$\frac{2}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	1
	$(\frac{12}{120})$	$(\frac{12}{120})$	$(\frac{24}{120})$	$(\frac{24}{120})$	$(\frac{24}{120})$	$(\frac{12}{120})$	$(\frac{12}{120})$	

一般に $n = n_1 + n_2$ の場合、 H_0 の下における $D = \bar{X} - \bar{Y}$ の確率分布は、 ${}_nC_{n_1}$ 通りの場合における D の値を計算することにより得られる。

帰無仮説 $H_0: F = G$ を, 対立仮説 $H_1: F > G$ に対して検定する場合には, 上で得られた D の確率分布において, 実際に観測された標本間の平均値の差 (D_0 とする) と等しいか, それより小さい D の個数を求めて K とすれば, 求める P -値は次式で与えられる。

$$\Pr \{D \leq D_0\} = K / {}_n C_{n_1} \quad \cdots \cdots \cdots \textcircled{1}$$

この P -値を, 0.05 とか 0.01 と較べて H_0 の棄却・採択をきめればよい。

さて本題にかえて

$x_1, \cdots, x_{n_1}; y_1, \cdots, y_{n_2}$ をプールして, この順に $xy(1), xy(2), \cdots, xy(n)$ とあらためて標記することにし,

$$D_{ij} = xy(i) - xy(j), (i, j = 1, \cdots, n) \quad \cdots \cdots \cdots \textcircled{2}$$

と定義すれば, ${}_n C_{n_1}$ 個の D は一般に

$$D = \bar{X} - \bar{Y} = \sum_{i \in (X)} \sum_{j \in (Y)} D_{ij} / (n_1 \cdot n_2) \quad \cdots \cdots \cdots \textcircled{3}$$

と表すことができる。ここに, $\sum_{i \in (X)} \sum_{j \in (Y)}$ は, i は X データに指定されたデータの全部, j は Y データに指定されたデータの全部にわたり, 和をとることを意味する。

X, Y 群へのデータの割り付け方と, H_0 の下における各場合の等確率性により, 次の関係式が成立する。

$$E(D | P, H_0) = 0, \quad \cdots \cdots \cdots \textcircled{4}$$

$$V(D | P, H_0) = \frac{n_1 \cdot n_2}{n(n-1)} \sum_{i=1}^n \left(\sum_{j=1}^n D_{ij} / (n_1 \cdot n_2) \right)^2 \quad \cdots \cdots \cdots \textcircled{5}$$

ここに, P は観測されたデータのパターンを表す。

$n \rightarrow \infty$ のとき, D の分布は, $N(0, V(D | P, H_0))$ に近づく。

なお, D の分布の対称性については, ties が存在したり, $n_1 \neq n_2$ の場合には, 保証されない上に, 少数例の場合には殊更に D の分布の離散度が大きいために, 漸近分布による P -値の近似計算は避けなければならない。

$G(t) = F(t - \Delta)$ が仮定できる場合には, 上述の仮説は, $H_0: \Delta = 0, H_1: \Delta > 0$ と表すことができ, 母集団分布のずれの検定ということができる。さらに $F(t)$ の正規性が仮定できれば, この Fisher の並べかえ検定は, すべての並べかえ検定の中で最も強力な検定であることが知られている。また, $F(t)$ が正規分布より裾の重い分布の場合には, 平均値の差より,

中央値の差を検定統計量とする方が、より強力であることも知られている。

この検定で最後に一つ注意しなければならないのは、観測値の中にど外れ値 (outliers) が含まれている場合である。outlier であるかないかの判別は簡単ではないが、医学データの中にはしばしば混入する心配がある。プロトコールに違反することが明らかであれば、除外して検定すればよいが、単に仲間から離れているから、棄却検定にかけたら棄却できるから、という理由で安直に除外するのは望ましいことではない。十分納得のいかない場合には、簡単に除外しないで、次の Wilcoxon の検定にかけるのも一法である。Wilcoxon の検定は、outliers の影響をあまり強く受けない検定法だからである。

ちなみに、ど外れ値が1個に含まれている場合には、それがX標本データに含まれるかY標本データに含まれるかにより、Dのヒストグラムは大きく左右に分離し、漸近正規分布とは似ても似つかぬ鬼っ子分布になるので、注意が肝要である。

A. (2) Wilcoxon の精密検定

Fisher の並べかえ検定と同じ考え方による検定法で、データの中にど外れ値 (outliers) の疑いがあるものが混入している場合などには、一層優れた安定した検定法ということが出来る。この検定では、データを生のまま用いないで、大きさの順位に直したものをを用いる。従って、平均値の差ではなく、両群データの平均順位の差を検定統計量とし、両母集団分布の中央値の差または、ずれの検定を目的としている。

$n = n_1 + n_2$ 個のデータ $x_1, x_2, \dots, x_{n_1}; y_1, y_2, \dots, y_{n_2}$ をプールして、大きさの順位を定め、この順に $r_1, r_2, \dots, r_{n_1}; s_1, s_2, \dots, s_{n_2}$ とする。このとき同順位データ (ties) があれば、中間順位を与える。 n 個の順位の列: $r_1, r_2, \dots, r_{n_1}; s_1, s_2, \dots, s_{n_2}$ を、この順にあらためて $rs(1), rs(2), \dots, rs(n)$ と標記する。この順位列のすべての順列を考え、初めの n_1 個をX標本の順位データ、残りの n_2 個をY標本の順位データと考えるのである。同順位がある場合にも、ごくわずかずつ異なると考え、 $n!$ 通りの順列について、それぞれ平均順位の差を求めるなど、A.(1)の検定法と全く同様である。

両母集団の中央値の差またはずれの検定における帰無仮説および対立仮説は、A(1)と同じく

$$H_0: F(t) = G(t), \quad H_1: F(t) > G(t)$$

とする。

また、A.(1)の D_{ij} と同様に

$$DR_{ij} = rs(i) - rs(j), (i, j = 1, 2, \dots, n) \dots\dots\dots ⑥$$

及び DR_{ij} にもとづく検定統計量 DR を

$$DR = \bar{R} - \bar{S} = \sum_{i \in (X)} \sum_{j \in (Y)} DR_{ij} / (n_1 \cdot n_2) \dots\dots\dots ⑦$$

と定義する。ここに $\sum_{i \in (X)} \sum_{j \in (Y)}$ は、 i は X データに指定されたデータの全部、 j は Y データに指定されたデータの全部にわたって和をとることを意味する。

DR を実際に計算するのは、 ${}_nC_{n_1}$ 通りの場合についてであり、帰無仮説 H_0 の下では、これらの各場合の起る確率は皆等しいから、DR の確率分布を容易に計算することができる。

観測された標本に対する DR の値を DR_0 とすれば、 H_0 を H_1 に対して検定する場合、求める P -値は、

$$\Pr \{DR \leq DR_0\} = K / {}_nC_{n_1} \quad \dots\dots\dots (8)$$

に等しい。ここに、 K は ${}_nC_{n_1}$ 個の DR 値の中で、 DR_0 に等しいか、それより小さいものの個数である。

H_0 の下における DR の分布については、A.(1)と同様に次式が成立する。

$$E(DR \mid P, H_0) = 0 \quad \dots\dots\dots (9)$$

$$V(DR \mid P, H_0) = \frac{n_1 \cdot n_2}{n(n-1)} \sum_{i=1}^n \left(\sum_{j=1}^n DR_{ij} / (n_1 \cdot n_2) \right)^2 \quad \dots\dots\dots (10)$$

ここに P は、観測された n 個のデータにもとづく順位パターンを表わす。

$n \rightarrow \infty$ のときには、DR の分布は $N(0, V(DR \mid P, H_0))$ に近づく。また、 n 個の観測データの中に ties があつたり、 $n_1 \neq n_2$ のときには、DR の分布の対称性が崩れるが、精密分布によっているので問題はない。

4. 検定モデルとその理論 (その2)

B. (1) 符号つき差の計による検定

医学・生物学の実験では、実験の目的としている処理要因の外に、実験結果に影響を及ぼす要因が無数にあり、しかもそれらを管理または制御できないことが多い。特に個体差にもとづく変動が大きいために、自然につくられた層またはブロックとして、一卵性双生児、同腹の仔、一枚の葉の左右半葉などの対 (pair) を取りあげ、これらを積極的に利用することが多い。ある薬剤の効果を、投与前と投与して一定時間後とにおける特性値の対観測値を比較することによって行おうとするのも、同じような考え方に沿うものである。こうした対応づけられた2標本にもとづく検定の問題を考察しよう。

これらの場合には、対応づけられた2つのデータ (対データ) の個々の値よりも、個体差にもとづく変動を除くため、それらの差に関心をあつめ、統計的推測のデータとする。このとき処理効果を示す差の有意性の検定において、差の母集団分布関数 $F(x - \theta)$ の連続性と、中央値 θ のまわりの対称性を仮定するが、この仮定は周到に計画された実験の対データの差に関し

ては満足されると見なしてよい場合が多い。

対応する2標本の対データの差を

$$X_i = X_{1i} - X_{2i}, \quad (i = 1, 2, \dots, n) \quad \dots\dots\dots(11)$$

とし、 X_1, X_2, \dots, X_n は、互いに独立で、 θ に関して対称で連続な分布関数 $F(x - \theta)$ に従うものとする。ここに X_{1i}, X_{2i} は、対観測値で、数値例でいえば、第 i 個体の薬剤注射直前の好酸球の率と、注射20分後の好酸球の率とである。

帰無仮説：「注射によって好酸球の百分率は変らない」は

$$H_0: \theta = 0$$

と表わすことができ、数値例の場合、薬の注射による好酸球の率の増加が期待されているから、対立仮説としては、

$$H_1: \theta < 0$$

を用いるのが適当であろう。

符号つき差の計による検定では、検定統計量 XWA として

$$XWA = X_1 + X_2 + \dots + X_n \quad \dots\dots\dots(12)$$

を用いる。

いま、 X_1, \dots, X_n の観測データを x_1, \dots, x_n としよう。このとき観測せられた標本データに対する検定統計量の値 (XWA_0 とする) は

$$XWA_0 = x_1 + x_2 + \dots + x_n$$

である。帰無仮説 H_0 の下では、差の母集団分布は原点に関して対称であるから

$$\Pr\{X_i = |x_i|\} = \Pr\{X_i = -|x_i|\} = 1/2 \quad \dots\dots\dots(13)$$

$$E(X_i | H_0) = 0 \quad \dots\dots\dots(14)$$

$$V(X_i | H_0) = x_i^2 \quad \dots\dots\dots(15)$$

が成立する。

以上により、 XWA は

$$\pm |x_1| \pm |x_2| \pm \dots \pm |x_n| \quad \dots\dots\dots(16)$$

において、 $+$ のすべての組合せからなる 2^n 個の値を、 H_0 の下では、すべて等確率 2^{-n} でとることが分る。これから H_0 の下における検定統計量 XWA の確率分布を定めることができる。こんどは、 n 次元空間の 2^n 個の点集合をベースに考えているわけである。 $|x_i|$ の中に

ties があるときも、ごくわずかず異なると考えて、検定統計量の確率分布を考えることも、並べかえ検定の場合と同様である。

式⑬～⑯より、XWA について

$$-\sum_{i=1}^n |x_i| \leq XWA \leq \sum_{i=1}^n |x_i| \quad \cdots \cdots \cdots \textcircled{17}$$

$$E(XWA | P, H_0) = 0 \quad \cdots \cdots \cdots \textcircled{18}$$

$$V(XWA | P, H_0) = \sum_{i=1}^n x_i^2 \quad \cdots \cdots \cdots \textcircled{19}$$

が成立する。ここにPは、対データの観測値の差のパターンを表わす。

$n \rightarrow \infty$ のとき、XWA が正規分布 $N(0, V(XWA | P, H_0))$ に近づくことは、⑯の形から類推されるように、中心極限定理の一般化により証明される。

B. (2) Wilcoxon の符号順位和精密検定

この検定は、対データ標本に対する Wilcoxon の符号順位和検定としてよく知られた検定の精密化と、ties による分布の乱れの克服を目指している。また、この検定のB.(1)に対する関係は、A.(2)のA.(1)に対する関係と非常によく似ている。

B.(1)の $X_i = X_{1i} - X_{2i}$ ($i = 1, \dots, n$) の実現値 $x_i = x_{1i} - x_{2i}$ ($i = 1, \dots, n$) から、 $|x_1|, |x_2|, \dots, |x_n|$ を導き、大きさによる順位を定めて、順に r_1, r_2, \dots, r_n とする。同順位データがあるときには中間順位を与える。 X_1, \dots, X_n は、B.(1)の場合と同じ理由で、差の母集団分布関数 $F(x - \theta)$ の連続性と、中央値 θ のまわりの対称性を仮定する。従って H_0 の下では、 R_1, R_2, \dots, R_n は、互いに独立な無作為標本で、 R_i は r_i と $-r_i$ の値をそれぞれ $1/2$ の確率でとる確率変数となる。

B.(1)のXWAに対応する検定統計量として

$$RWA = R_1 + R_2 + \dots + R_n \quad \cdots \cdots \cdots \textcircled{20}$$

を用い、帰無仮説 $H_0: \theta = 0$ を対立仮説 $H_1: \theta < 0$ に対して検定する問題を考察しよう。

観測された対標本におけるRWAの値を RWA_0 とすると

$$RWA_0 = \sum_{i=1}^n \text{sign}(x_i) \cdot r_i$$

次に、 H_0 の下におけるRWAの値は

$$\pm r_1 \pm r_2 \pm \dots \pm r_n \quad \cdots \cdots \cdots \textcircled{21}$$

において、+-のすべての組合せによる 2^n 個の値を、それぞれ 2^{-n} の確率でとることが分かる。従って、RWA の 2^n 個の値のうち、 RWA_0 に等しいか、それより小さい値の個数を

Kとすると、観測された対標本の H_0 の下での P-値は

$$\Pr \{RWA \leq RWA_0\} = K \cdot 2^{-n} \quad \dots\dots\dots (22)$$

で与えられる。

また、次の関係式が成立する。

$$-\sum_{i=1}^n r_i \leq RWA \leq \sum_{i=1}^n r_i \quad \dots\dots\dots (23)$$

$$E(RWA | P, H_0) = 0 \quad \dots\dots\dots (24)$$

$$V(RWA | P, H_0) = \sum_{i=1}^n r_i^2 \quad \dots\dots\dots (25)$$

ここに、Pは、対データの観測値の差の、大きさの順位のパターンを表わす。

$n \rightarrow \infty$ のとき、RWA の分布は $N(0, \sum_{i=1}^n r_i^2)$ に近づく。このことは、B.(1)と同様、中心極限定理の一般化により証明される。

5. BASIC プログラムと出力の一部

A.(1)のプログラムと、B.(2)のプログラムを掲げ、それらをそれぞれ数値例に適用した場合の出力の一部を次に示そう。

A.(1)とA.(2)とは、 $n_1 + n_2 = n$ 個のデータを、すべての可能な方法でX群、Y群に割り付ける手順において同一であり、B.(1)とB.(2)とは、 n 個の値の+のすべての組合せを作る手順において同じである。またA.(2)とB.(2)とは、観測されたデータを順位データに変換するプログラムにおいて共通である。そこでA.(1)とB.(2)のプログラムを参考にすれば、A.(2)、B.(1)のプログラムは容易に組めるはずである。

$n = n_1 + n_2$ 個のデータにもとづいて、すべての可能なX群、Y群への分割を行い、平均値の差を計算するのは、表3の1730行~2200行で、このプログラムの核心である。与えられた任意の n と n_1 に対応できるところがポイントである。 D_{ij} ($i, j = 1, 2, \dots, n$) の $n \times n$ 行列を計算しておけば、④、⑤により、それぞれのDの計算に先立って、Dの平均(理論的に0であることが分かっているが、プログラム上で確認することは必要である)と標準偏差SDDが算定できるので、Dの度数分布表の作成に甚だ好都合である。Dの確率分布のグラフ表示には、 n, n_1 およびデータの多様性に対応するため工夫が必要であるが、検定の実際には直接関係はない。しかし漸近分布によるP-値と、精密に計算されたP-値の関係について理解を深め、納得するためには、やはりグラフ表示を見送るわけにはいかないであろう。1720行のif文を無条件GOTO文に換えれば、 n の大きさに拘らず、直ちに漸近分布にもとづくP-値が得られる。Aの数値列では $n = 5 + 6 = 11$ であるが、1720行をGOTO 2860としてP-値(正規近似値)の算定もしてある。

“,” による注釈行は, いわばプログラム作成の足場みたいなもので, なくもがなの面もあるが, 検定手順を詳細にチェックしたいときには, 簡単に“,”を DELETE すれば, リストが出力されるので便利であろうと考えて残しておいた。

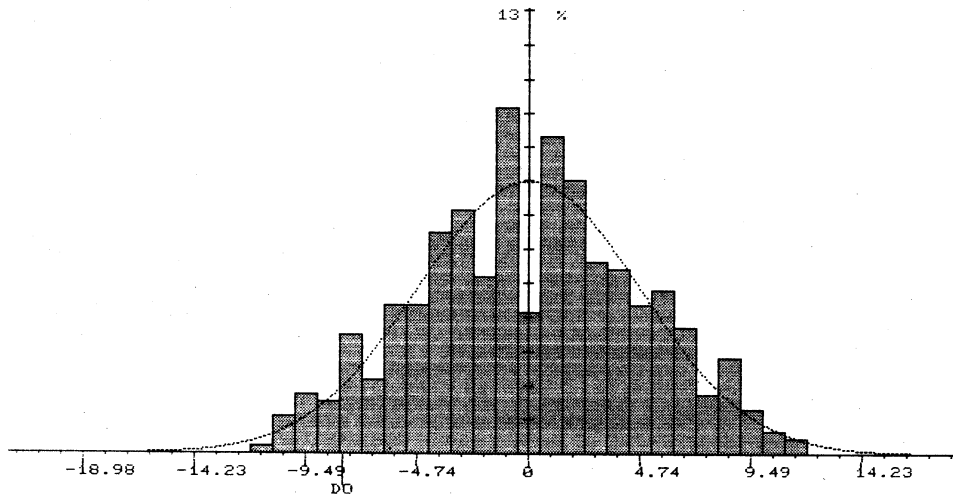
A.(1)を数値例に適用した結果の出力の一部を図1に示した。図1を仔細に検討すれば ties によるDの確率分布の正規性からの崩れ方や, P-値の精密な値と漸近分布による近似値との大小関係も納得がいくであろう。この数値例では, 漸近分布近似によるP-値は0.0473であるから, 5%水準で H_0 を H_1 に対し棄却することが出来る。しかし精密計算によればP-値は0.0519となるから, 棄却することは出来ない。

A.(2), B.(2)では, まず観察されたデータの差を, 大きさの順に並べ, 順位データに変換しなければならないが, これはB.(2)のプログラム(表4)のサブルーチン***RANKING**を用いて行われる。このさい tie の組数やその長さを調べて中間順位を決めなければならない。B.(2)のプログラムでは, 入力データとその順位データの関係が分り易いように, サブルーチン***QUICKS** (Quick Sort)を用いて, 与えられた順にデータを並べなおしている。(1360行~1400行)。この***RANKING**と***QUICKS**とをうまく活用すれば, A.(1)のプログラムをA.(2)に組みかえることはそれほど困難ではあるまい。もっともA.(2)のDRは整数値をとるので, その特長を活かそうとすれば, 度数分布表やグラフ表示に一工夫が必要である。

次にB.(2)のプログラム(表4)では, $\pm r_1 \pm r_2 \pm \dots \pm r_n$ において, +のすべての組合せに対するRWAの算定が必要であるが, これは1490行~1520行で行われている。+, -は, 2進表示の0, 1に対応づけられている。RWAの値は整数であるから, 度数分布表は階級に分けず, そのままの値による度数で表すことにした。このさい, INTによる配列番号Jの指定では, 例えば, -0と+0とが同じJになるよう工夫してある(1600行)。度数分布のグラフ表示においても, RWAの刻みが必ずしも一定ではない点を考慮したり, nが小さいときの柱の色合いについても少し工夫がしてある。

B.(2)を数値例に適用した場合の出力の一部を図2に示す。RWAの値は, nが15とかなり大きいので, ほぼ正規分布に従っているようである。しかし, RWAは分布の大部分で連続した整数値をとっているのに, 分布の両端付近では2とばしの整数値をとっている。また, 分布の中央付近はやや正規分布より低目になっており, P-値は正規近似によるとき, 0.0043, 精密計算によるとき0.0028であった。

Fisher's Permutation Exact Test

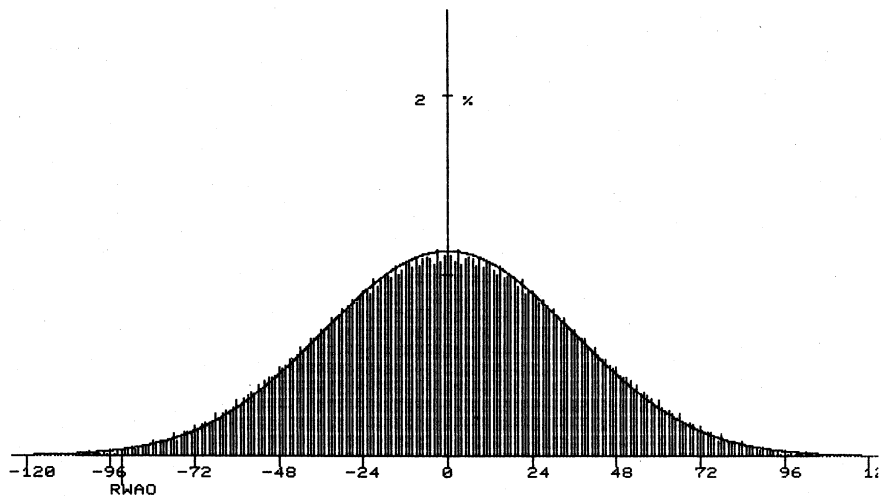


Exact P-value= 0.0519 (One-Sided)

Approximate Normal P-Value= 0.0473 (One-Sided)

図1 A.(1)を数値例に適用して得られたグラフ

Wilcoxon Signed Rank Exact Test



Exact P-VALUE= 0.0028 (One-Sided)

Approximate Normal P-VALUE= 0.0043 (One-Sided)

図2 B.(2)を数値例に適用して得られたグラフ

5. 結 語

少数例で、しかも同順位データの存在するいろいろの場合について、検定統計量の精密分布を求めて検討した結果、有意確率を漸近分布を用いて計算する現行の検定手順には、不都合な点が存在することが判明した。

研究者は、自分で実験データの処理を体験しながら、統計的検定の本当の意味を理解し、研究を推進することが必要であるが、パーソナルコンピューターが普及した今日では、それが可能になったと言うことが出来よう。

この論文にのせた統計的検定は、母集団分布について強い仮定を置かない検定法で、理論的にも優れており、直観的にも理解し易いものであるから、実際にコンピュータを走らせて、出力を辿ってゆけば、実験データを通じて、現象のより深い理解に達することができるであろう。

参 考 文 献

- [1] 仮谷太一 (1986) : A generalized Wilcoxon test for comparing interval-censored data samples. 第14回日本行動計量学会予稿集
- [2] 柳川 堯 (1982) : ノンパラメトリック法, 培風館

表3 A.(1): Fisher の並べかえ精密検定

```

1000 REM*****
1010 REM*
1020 REM* 独立2標本(計量データ)にもとづく
1030 REM* 平均値の差の検定・ずれの検定
1040 REM*
1050 REM* ----- 並べかえによる精密検定 -----
1060 REM*
1070 REM* 1986.6.4. By T.Kariya
1080 REM*
1090 REM*****
1100 REM Fisher's Permutation Exact Test. File Name:"FPET-I2S"
1110 REM Ties(Tied Data) are not prohibited.
1120 REM Ho: F=G, H1: F>G, H2: F<G, H3: F>G or F<G
1130 REM G(t)=F(t-d) または処理効果の加法性がなりたち、しかも
1140 REM t(t)が正規分布のとき、並びかえ検定のなかで最強力である
1150 REM DATA 1:NX,NY,N (Data Size, N=NX+NY)
1160 REM DATA 2:X1,X2,X3,...,Xnx.
1170 REM DATA 3:Y1,Y2,Y3,...,Yny.
1180 DEFINT R
1190 DATA 5.6,11: 'データ数
1200 DATA 31,28,19,31,33: 'Xデータ
1210 DATA 40,33,47,23,37,38: 'Yデータ
1220 HAJIME$=TIME$
1230 READ NX,NY,N
1240 NXY=NX*NY:NXY2=2*NXY
1250 DIM JX(NX),JY(NY),XY(N),DIJ(N,N),GD(SO)
1260 FOR I=1 TO N:READ XY(I):NEXT I: 'データの読み込み
1270 LPRINT TAB(5):"並べかえによる平均値の差 または ずれの精密検定"
1280 LPRINT:LPRINT:PRINT STRING$(SS,"-")
1290 LPRINT USING"NX=####":NX: '----- 入力データの表示 -----
1300 FOR I=1 TO NX
1310 LPRINT USING"####.##":XY(I):
1320 NEXT I
1330 LPRINT:LPRINT USING"NY=####":NY
1340 FOR I=NX+1 TO N
1350 LPRINT USING"####.##":XY(I):
1360 NEXT I
1370 LPRINT:PRINT STRING$(SS,"-"):LPRINT
1380 SYMBOLQ(350,250):"計算中",3,3,3
1390 REM -----DIJ=Xi-Yj の表 -----
1400 SXY=0:FOR J=1 TO N:SXY=SXY+XY(J):NEXT J
1410 FOR K=1 TO N-1
1420 FOR KK=K+1 TO N
1430 DIJ(K,KK)=XY(K)-XY(KK)
1440 NEXT KK,K
1450 FOR I=1 TO N:DIJ(I,I)=0:NEXT I
1460 FOR K=2 TO N
1470 FOR KK=1 TO K-1
1480 DIJ(K,KK)=-DIJ(KK,K)
1490 NEXT KK,K
1500 DO=0
1510 FOR I=1 TO NX
1520 FOR J=NX+1 TO N
1530 DO=DO+DIJ(I,J)
1540 NEXT J,I
1550 DO=DO/NXY
1560 SDI=0:SDI2=0
1570 FOR I=1 TO N
1580 DI=0
1590 FOR J=1 TO N
1600 DI=DI+DIJ(I,J)
1610 NEXT J
1620 SDI=SDI+DI:SDI2=SDI2+DI*DI
1630 NEXT I
1640 'LPRINT:LPRINT:PRINT " DIJ=Xi-Yj の リスト "
1650 FOR K=1 TO N
1660 'LPRINT
1670 FOR KK=1 TO N
1680 'LPRINT USING"####.##":DIJ(K,KK):
1690 NEXT KK,K
1700 VD=SDI2*NXY/(N*(N-1)):SDD=SQR(VD)
1710 TDM=SDI/NXY:TVD=SDI2/NXY/(N*(N-1)):TSDD=SQR(TVD)
1720 IF N>18 GOTO 2860: ' --- 大標本ならば漸近正規分布へと ---
1730 REM ***** 小標本の場合 *****
1740 DHABA=INT(SDD*100)/500
1750 'LPRINT:LPRINT " Data Number"
1760 'LPRINT " NO of X-Group D ":LPRINT
1770 R=NX:CN=0:NMR=N-R:LPRINT
1780 I=1:JX(1)=1:SD=0:VD=0:PNC=0
1790 IF I=R GOTO 1840
1800 IPI=I+1
1810 FOR L=IPI TO R
1820 JX(L)=JX(L-1)+1
1830 NEXT L
1840 CN=CN+1
1850 REM ----- データを 2つの群に 分割する -----
1860 I=1:K=1:II=0
1870 IF I<>JX(K) GOTO 1900
1880 IF I<N THEN I=I+1:GOTO 1870
1890 GOTO 1940
1900 IF I=JX(K) GOTO 1920
1910 IF K<R THEN K=K+1:GOTO 1870
1920 II=II+1:JY(II)=I
1930 IF I<N THEN I=I+1:GOTO 1870
1940 'LPRINT USING"#####":CN:
1950 FOR KX=1 TO R
1960 'LPRINT USING"#####":JX(KX):
1970 NEXT KX
1980 '--- FOR KY=1 TO NMR:LPRINT USING"#####":JY(KY):NEXT KY
1990 REM ----- 平均値の差:D=XBAR-YBAR の 計算 -----
2000 D=0
2010 FOR K=1 TO R
2020 FOR KK=1 TO N-R
2030 D=D+DIJ(JX(K),JY(KK))
2040 NEXT KK,K
2050 TD=D/NXY
2060 J=INT(D/DHABA+.5)+25
2070 GD(J)=GD(J)+1
2080 IF DO<0 GOTO 2110
2090 IF D<DO GOTO 2190
2100 PNC=PNC+1:GOTO 2190
2110 IF D>DO GOTO 2130
2120 PNC=PNC+1
2130 'LPRINT USING" #####.####":TD
2140 REM -----
2150 I=R
2160 IF JX(I)<NMR+I GOTO 2190
2170 I=I-1
2180 IF I<0 THEN 2210 ELSE 2160
2190 JX(I)=JX(I)+1
2200 GOTO 1790
2210 LPRINT:LPRINT USING"E(DIP,H)=###.##, V(DIP,H)=#####.##, SD(DIP,H)=#####.###":TDM,TVD,TSDD
2220 LPRINT:LPRINT USING"NO. of all cases=#####":CN
2230 LPRINT:LPRINT USING"NO. of more discrepant D than DO, K=#####":PNC
2240 LPRINT:LPRINT USING"D(Observed)=###.####, (Z(Observed)=#####)":TD,DO/SDO
2250 LPRINT:LPRINT USING"Exact P-value=###.#### (One-Sided)":PNC/CN
2260 REM ----- D の 度数分布表 -----
2270 SGD=0:LPRINT:LPRINT
2280 LPRINT:LPRINT " FREQUENCY TABLE OF D"
2290 LPRINT:LPRINT " CLASS FREQ. PROBABILITY(X)":LPRINT
2300 FOR I=0 TO 2*25
2310 IF GD(I)=0 GOTO 2350
2320 CMN=(I-25)*DHABA/NXY-DHABA/2/NXY
2330 LPRINT USING" ###.## -- ###.## #####":CMN,CMN+DHABA/NXY,GD(I),GD(I)/CN*100
2340 SGD=SGD+GD(I)
2350 NEXT I
2360 LPRINT USING" TOTAL #####.####":SGD,SGD*100/CN
2370 LPRINT:LPRINT USING"NO. OF ALL CASES=#####":CN
2380 CLS: '----- 度数分布図 (ヒストグラム) -----
2390 BDI=2.5066*SDD/DHABA: '2.5066=SQR(2*PAI)
2400 PMAX=0
2410 FOR I=0 TO 2*25
2420 IF PMAX<GD(I) THEN PMAX=GD(I)
2430 NEXT I
2440 PMAX=INT(PMAX*100/CN)+1
2450 JHABA=INT(260/PMAX)
2460 IHABA=15
2470 FOR ID=-25 TO 25
2480 IID=ID+25
2490 PRO=GD(IID)*100/CN
2500 X=350+IHABA*ID:Y=350-JHABA*PRO
2510 LINE(X-IHABA/2,350)-(X+IHABA/2,Y),PSET,7,B
2520 PAINT (X,350-JHABA*PRO/2),&H550055AAAA00,7
2530 LINE(X,350)-(X,352),PSET,5
2540 REM ----- 検定値の目盛 -----
2550 CDG=INT(ABS(ID)*DHABA/NXY*100)/100:CDG=SGN(ID)*CDG:Z$=STR$(CDG)
2560 IF (ID MOD 5)<>0 GOTO 2590
2570 LINE (X,350)-(X,357),PSET,5
2580 SYMBOL(X-12,358),Z$.1,1,5
2590 NEXT ID
2600 LINE (350,50)-(350,350),PSET,7
2610 LINE (100,350)-(600,350),PSET,7
2620 REM ----- 縦座標の目盛 -----
2630 FOR J=0 TO 20
2640 Y=350-JHABA*J
2650 IF Y<50 GOTO 2680
2660 LINE (347,Y)-(353,Y),PSET,7
2670 NEXT J
2680 Z$=STR$(J-1)+"" 'Y=350-JHABA*(J-1)
2690 SYMBOL (320,Y),Z$.1,1,6
2700 REM --- D の 漸近分布は 正規分布 N(0,VD) -----
2710 FOR ID=-25 TO 25 STEP .1
2720 Z=ID*DHABA/SDO
2730 SFZ=EXP(-Z*Z/2)*100/BDI
2740 X=350+IHABA*ID:Y=350-JHABA*SFZ
2750 PSET (X,Y),7
2760 NEXT ID
2770 X=350+IHABA*DO/DHABA
2780 LINE(X,350)-(X,370),PSET,6
2790 SYMBOL(X-8,370),DO".1,1,6
2800 SYMBOL (100,25),"Fisher's Permutation Exact Test",2,2,6
2810 LPRINT:LPRINT:LPRINT:LPRINT:LPRINT
2820 HARDC1
2830 LPRINT:LPRINT "START=":HAJIME$,
2840 LPRINT "END=":TIME$
2850 STOP
2860 REM ***** 大標本の場合 *****
2870 LPRINT:LPRINT:LPRINT USING"V(DIP,H)=#####.##, SD(DIP,H)=#####.####":TVD,TSDD
2880 ZO=DO/SDO
2890 IF ZO>0 THEN ZN=-ZO ELSE ZN=ZO
2900 GOSUB *NLO
2910 LPRINT:LPRINT USING"D(Observed)=###.####, Z(Observed)=#####":TDO,ZO
2920 LPRINT:LPRINT USING"Approximate Normal P-Value=###.#### (One-Sided)":Z
2930 STOP
2940 *NLO
2950 A1H=.0705230784H:A2H=.0422820123H:A3H=.0092705272H
2960 A4H=.0001520143H:A5H=.0002765672H:A6H=.0000430638H
2970 IF ZN<-5.5 GOTO 3040
2980 ZH=.707106781H*ABS(ZN)
2990 FZH=1H+(A1H+(A2H+(A3H+(A4H+(A5H+A6H*ZH)*ZH)*ZH)*ZH)*ZH
3000 IF FZH=10000H GOTO 3040
3010 FZH=FZH*FZH:FZH=FZH*FZH:FZH=FZH*FZH:FZH=FZH*FZH
3020 FZH=1H-FZH/FZH
3030 GOTO 3050
3040 FZH=1H
3050 FZH=.5H*(1H-FZH)
3060 Z=FZH
3070 RETURN

```


表4 B.(2): Wilcoxon の符号順位と精密検定

```

1000 REM*****
1010 REM*
1020 REM*      ▲      対応2標本の 対データの差 にもとづく
1030 REM*
1040 REM*      中央値または ずれの 符号順位と精密検定
1050 REM*
1060 REM*      1986.6.5.      T.KARIYA
1070 REM*
1080 REM*****
1090 REM 差の母集団分布の連続性と 中央値のまわりの対称性を仮定する
1100 REM Generalized Wilcoxon Signed Rank Exact Test:"W-SRET"
1110 REM DATA 処理前の値,処理後の値
1120 DATA 15:'対データの数
1130 DATA 2.0.3.4 : '対データ その1
1140 DATA 10.2.15.6: '対データ その2
1150 DATA 6.0.5.4 : '対データ その3
1160 DATA 7.8.10.8 : '対データ その4
1170 DATA 2.2.5.6 : '対データ その5
1180 DATA 3.4.6.2 : '対データ その6
1190 DATA 1.6.3.8 : '対データ その7
1200 DATA 4.2.3.4 : '対データ その8
1210 DATA 6.6.5.2 : '対データ その9
1220 DATA 7.6.10.0 : '対データ その10
1230 DATA 3.8.2.0 : '対データ その11
1240 DATA 8.6.13.2 : '対データ その12
1250 DATA 4.2.5.4 : '対データ その13
1260 DATA 11.2.13.6: '対データ その14
1270 DATA 2.0.4.4 : '対データ その15
1280 HAJIME$=TIME$
1290 READ N:NGW=N*(N+1)/2:N2=2^N: NGW=MAX(RWA)
1300 DIM X(N,2),SAC(N),SAA(N),A(N),B(N),RANK(N),NO(N),NI(N),GW(2*NGW+1),TAU(N)
1310 LPRINT TAB(5):"一般化 Wilcoxon 符号順位と検定":LPRINT
1320 GOSUB *READDATA:'データの 入力
1330 GOSUB *SUBTRACT:'(第1群データ)-(第2群データ)
1340 FOR I=1 TO N:NO(I)=I:NEXT I
1350 GOSUB *RANKING:'差の絶対値による順位付け (Tiesがあってもよい)
1360 REM ---- 与えられたデータ順に並べなおす ----
1370 FOR I=1 TO N:SWAP A(I),B(I):NEXT I
1380 IW=1:JW=N:K=0
1390 IF IW<JW THEN GOSUB *QUICKS:K=K+1:JS(K)=JW:IS(K)=MW:JW=MW-1:GOTO 1390
1400 IF K>0 THEN MW=IS(K):JW=JS(K):K=K-1:IW=MW+1:GOTO 1390
1410 GOSUB *PRINTDATA:'入力データの表示
1420 SYMBOLQ(350,250):"計算中",3.3,3
1430 VRWA=0:'RWA=Sum of Signed Rank
1440 RWAO=0:FOR I=1 TO N:RWAO=RWAO+SGN(SA(I))*RANK(I):NEXT I
1450 FOR I=1 TO N:VRWA=VRWA+RANK(I)^2:NEXT I
1460 SDRWA=SQR(VRWA)
1470 IF N>18 GOTO 2400:'大標本のときはとべ
1480 REM ***** 少数例の場合 *****
1490 REM ---- 仮説: H0のもとで 可能なすべての場合のRWAを計算する ----
1500 CN=0:PNC=0:SRWA=0
1510 FOR I=0 TO N-1:NI(I)=0:NEXT I
1520 RWA=0:FOR I=0 TO N-1:RWA=RWA+(-1)^NI(I)*RANK(I+1):NEXT I
1530 SRWA=SRWA+RWA
1540 IF RWAO<0 GOTO 1570
1550 IF RWA<RWAO GOTO 1590
1560 PNC=PNC+1:GOTO 1590
1570 IF RWA>RWAO GOTO 1590
1580 PNC=PNC+1
1590 CN=CN+1
1600 J=INT(RWA+.000001):GW(J+NGW)=GW(J+NGW)+1
1610 FOR K=0 TO N-1
1620 IF NI(K)=0 THEN NI(K)=1:GOTO 1520
1630 NI(K)=0
1640 NEXT K
1650 LPRINT
1660 REM ----- 度数分布表 -----
1670 SGW=0:LPRINT:LPRINT
1680 LPRINT:LPRINT" FREQUENCY TABLE OF RWA"
1690 LPRINT:LPRINT" VALUE FREQ. PROBABILITY(%):":LPRINT
1700 FOR I=-NGW TO NGW
1710 IF GW(I+NGW)=0 GOTO 1740:' 度数が 0 なら とべ
1720 LPRINT USING"#####.H ... #####.#####":I,GW(I+NGW),GW(I+NGW)/N2*100
1730 SGW=SGW+GW(I+NGW)
1740 NEXT I
1750 LPRINT USING" TOTAL #####.#####":SGW,SGW*100/N2
1760 LPRINT:LPRINT USING"NO. OF ALL CASES=#####":CN:LPRINT
1770 LPRINT USING"NO. OF MORE DISCREPANT RWA THAN RWAO. K=#####":PNC
1780 ERWA=SRWA/CN:LPRINT
1790 LPRINT USING"E(RWA|P,H)=#####.H, V(RWA|P,H)=#####.H, SDRWA=#####.H,ERWA,VRW
A,SDRWA:LPRINT
1800 LPRINT USING"RWAO=#####.H, Exact P-VALUE=#####.H(One-Sided)":RWAO,PNC/N2
1810 CLS:'----- 度数分布図 (ヒストグラム) -----
1820 B01=2.5066*SDRWA:'2.5066=SQR(2*PAI)
1830 PMAX=0

```

```

1840 FOR I=0 TO 2*NGW
1850 IF PMAX<GW(I) THEN PMAX=GW(I)
1860 NEXT I
1870 PMAX=PMAX*100/N2
1880 JHABA=INT(260/(PMAX+1))
1890 IHABA=INT(250/3)
1900 REM ----- 横座標の目盛 -----
1910 MR=INT(NGW/5)
1920 FOR J=-NGW TO NGW
1930 X=350+IHABA*J/SDRWA:Y=350
1940 IF (J MOD MR)<>0 GOTO 1960
1950 LINE(X,Y)-(X,Y+7),PSET,5
1960 NEXT J
1970 FOR J=0 TO NGW STEP MR
1980 X=350+IHABA*J/SDRWA:Y=350
1990 Z$=STR$(J)
2000 LINE(X,Y)-(X,Y+7),PSET,5
2010 SYMBOL(X-12,358),Z$,1.1,1.5
2020 X=350-IHABA*J/SDRWA:Y=350
2030 Z$=STR$(-J)
2040 LINE(X,Y)-(X,Y+7),PSET,5
2050 SYMBOL(X-12,358),Z$,1.1,1.5
2060 NEXT J
2070 XD=INT(250/NGW/3)
2080 FOR J=-NGW TO NGW
2090 JNGW=J+NGW
2100 PRO=GW(JNGW)*100/N2
2110 X=350+IHABA*J/SDRWA:Y=350-JHABA*PRO
2120 LINE(X-XD,350)-(X+XD,Y),PSET,7,8
2130 PAINT(X,350-JHABA*PRO/2),&H550055AAAA00,7
2140 NEXT J
2150 REM ----- 縦座標の目盛 -----
2160 LINE (350,50)-(350,350),PSET,7
2170 LINE (100,350)-(600,350),PSET,7
2180 FOR J=0 TO 30
2190 Y=350-JHABA*J
2200 IF Y<50 GOTO 2230
2210 LINE (347,Y)-(353,Y),PSET,7
2220 NEXT J
2230 Z$=STR$(J-1)+"" Y:Y=350-JHABA*(J-1)
2240 SYMBOL (320,Y),Z$,1.1,1.6
2250 REM --- RWA の漸近分布は 正規分布 N(0,VRWA) -----
2260 FOR IW=-NGW-4 TO NGW+4 STEP 1
2270 Z=IW/SDRWA
2280 SFZ=EXP(-Z*Z/2)*100/B01
2290 X=350+IHABA*Z:Y=350-JHABA*SFZ
2300 PSET (X,Y),7
2310 NEXT IW
2320 X=350+IHABA*RWAO/SDRWA
2330 LINE(X,350)-(X,370),PSET,6
2340 SYMBOL(X-8,370),"RWAO",1.1,1.6
2350 SYMBOL (100,25),"Wilcoxon Signed Rank Exact Test",2.2,6
2360 HARDCL
2370 LPRINT:LPRINT "START=":HAJIME$,
2380 LPRINT "END=":TIME$
2390 STOP
2400 REM ***** 大標本の場合 *****
2410 VU=RWAO/SDRWA
2420 ZN=(ABS(RWAO)-.5)/SDRWA
2430 GOSUB *NTP:'正規分布の左側または右側確率
2440 LPRINT:LPRINT USING"V(RWA|P,H)=#####.H, SD(RWA|P,H)=#####.H,VRW
A,SDRWA
2450 LPRINT:LPRINT USING"RWA(Observed)=###.####, Z(Observed)=###.####":RWAO,VU
2460 LPRINT:LPRINT USING"Normal Approximate P-VALUE=###.#### (One-Sided)":PRO
2470 CLS
2480 LPRINT:LPRINT "START=":HAJIME$,
2490 LPRINT "END=":TIME$
2500 STOP
2510 *NTP:'正規分布の左側または右側確率 サブルーチン
2520 A1H=.0705230784H:A2H=.0422820123H:A3H=.0092705272H
2530 A4H=.0001520143H:A5H=.0002765672H:A6H=.0000430638H
2540 ZH=.707106781H*ABS(ZN)
2550 IF ZH>10H THEN PRO=0:GOTO 2600
2560 FZH=1H+(A1H+(A2H+(A3H+(A4H+(A5H+A6H*ZH)*ZH)*ZH)*ZH)*ZH
2570 IF FZH=1000000000H THEN PRO=0:GOTO 2600
2580 FZH=FZH*FZH:FZH=FZH*FZH:FZH=FZH*FZH:FZH=FZH*FZH
2590 PRO=.5H/FZH
2600 RETURN
2610 *READDATA:'データの 入力 (サブルーチン)
2620 FOR I=1 TO N
2630 FOR J=1 TO 2
2640 READ X(I,J)
2650 NEXT J
2660 NEXT I
2670 RETURN

```

```

2680 *PRINTDATA:'入力データの表示 (サブルーチン)
2690 FOR I=1 TO N:SAA(I)=B(I):NEXT I
2700 LP=S
2710 FOR I=1 TO N STEP LP
2720 LPRINT " DATA"
2730 LPRINT STRING$(S6,"-")
2740 IE=I+LP-1:IF IE>N THEN IE=N
2750 LPRINT " 処理前 ":
2760 FOR II=I TO IE
2770 LPRINT USING" #####.H ":X(II,1):
2780 NEXT II
2790 LPRINT
2800 LPRINT " 処理後 ":
2810 FOR II=I TO IE
2820 LPRINT USING" #####.H ":X(II,2):
2830 NEXT II
2840 LPRINT:LPRINT STRING$(S6,"-")
2850 LPRINT " 01f. ":
2860 FOR II=I TO IE
2870 LPRINT USING" ###.####":SA(II):
2880 NEXT II
2890 LPRINT
2900 LPRINT " Rank ":
2910 FOR II=I TO IE
2920 LPRINT USING" ###.H ":RANK(II):
2930 NEXT II
2940 LPRINT:LPRINT STRING$(S6,"-"):LPRINT
2950 NEXT I
2960 RETURN
2970 *SUBTRACT:'差 (サブルーチン)
2980 FOR I=1 TO N
2990 SA(I)=X(I,1)-X(I,2)
3000 NEXT I
3010 RETURN
3020 *RANKING:'差の絶対値による順位付け (サブルーチン)
3030 FOR I=1 TO N:SAA(I)=ABS(SA(I)):A(I)=SAA(I):B(I)=NO(I):NEXT I
3040 IW=1:JW=N:K=0
3050 IF IW<JW THEN GOSUB *QUICKS:K=K+1:JS(K)=JW:IS(K)=MW:JW=MW-1:GOTO 3050
3060 IF K>0 THEN MW=IS(K):JW=JS(K):K=K-1:IW=MW+1:GOTO 3050
3070 IW=1:JW=0
3080 MRS=I:NER=1
3090 FOR J=I+1 TO N
3100 IF ABS(A(I)-A(J))>.000001 GOTO 3130
3110 MRS=MRS+J:NER=NER+1
3120 NEXT J
3130 RM=MRS/NER:K=I+NER-1
3140 IF NER>1 THEN JJ=JJ+1:TAU(JJ)=NER
3150 FOR II=I TO K
3160 RANK(II)=RM
3170 NEXT II
3180 I=K+1:IF I<N GOTO 3080
3190 IF I=N THEN RANK(I)=N
3200 RETURN
3210 *QUICKS
3220 IL=IW+1:IR=JW:AW=A(IW)
3230 FOR I=IL TO IR
3240 IF A(I)>AW THEN 3260
3250 NEXT I
3260 FOR J=IR TO IL STEP -1
3270 IF A(J)<=AW THEN 3290
3280 NEXT J
3290 IF I<J THEN 3300 ELSE 3320
3300 SWAP A(I),A(J):SWAP B(I),B(J):SWAP RANK(I),RANK(J)
3310 IL=I:IR=J:GOTO 3230
3320 SWAP A(J),A(IW):SWAP B(J),B(IW):SWAP RANK(J),RANK(IW)
3330 MW=J
3340 RETURN

```